

Report on QPTDat Project Workshop “FAIR Research Data in Plasma Medicine”

Markus M. Becker¹, Fabian Hoppe², Mathias Röper³, Tabea Tietz², Simon Tschirner³, and Laura Vilardell Scholten¹

¹Leibniz Institute for Plasma Science and Technology (INP), 17489 Greifswald, Germany

²FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH, 76344 Eggenstein-Leopoldshafen Germany

³Hochschule für Angewandte Wissenschaften Hamburg, 20099 Hamburg, Germany

Online Workshop, October 28-29, 2020

SPONSORED BY THE

The workshop was funded by the German Federal Ministry of Education and Research under grants no. 16QK03A, 16QK03B, and 16QK03C. The responsibility for the content of this publication lies with the authors.



Federal Ministry
of Education
and Research



The content of this publication is subject to the "[Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)" license.

DOI: [10.5281/zenodo.4350287](https://doi.org/10.5281/zenodo.4350287)

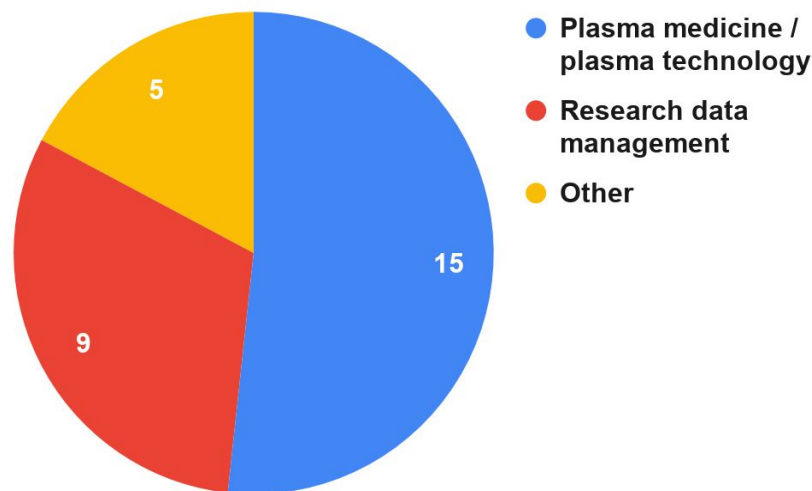
Participants

(List of attendees who agreed to be listed in the list of participants.)

	Name	Institution	Country
1	Markus Becker	INP Greifswald	Germany
2	Tabea Tietz	FIZ Karlsruhe	Germany
3	Mathias Röper	HAW Hamburg	Germany
4	Simon Tschirner	HAW Hamburg	Germany
5	Harald Sack	FIZ Karlsruhe	Germany
6	Zeinab Kabarkouhi	Beheshti University	Iran
7	Nick Plathe	INP Greifswald	Germany
8	Fabian Hoppe	FIZ Karlsruhe	Germany
9	Holger Israel	Physikalisch-Technische Bundesanstalt (PTB)	Germany
10	Junaid Khan	Comsats University Islamabad	Pakistan
11	Joachim Meier	Physikalisch-Technische Bundesanstalt (PTB)	Germany
12	Peter Hill	University of York	United Kingdom
13	Markus Niebel	Fraunhofer-Institute for Mechanics of Materials IWM	Germany
14	Marjan Stankov	INP Greifswald	Germany
15	Andrei Vasile Nastuta	'Gr. T. Popa' University of Medicine and Pharmacy Iasi	Romania
16	Robert Wagner	INP Greifswald	Germany
17	Marina Prenzel	Ruhr-Universität Bochum	Germany
18	Laura Vilardell	INP Greifswald	Germany
19	Sybille Hasse	INP Greifswald	Germany
20	Sepideh Mousazadeh Borghei	INP Greifswald	Germany
21	Rene Hackl-Sommer	FIZ Karlsruhe	Germany

22	Milad Rasouli	Kharazmi University	Iran
23	Frank Krüger	Universität Rostock	Germany
24	Soeun Paek	Sapienza University of Roma	Republic of Korea
25	Anke Schmidt	INP Greifswald	Germany
26	Robert Ott	IIB e.V. Warnemünde	Germany

Poll on the background of the participants:



Workshop goals

- Exchange about needs and benefits of making research data findable, accessible interoperable, and reusable (FAIR) involving both experts in the plasma domain and research data management (RDM) experts
- Identification of requirements and possibilities of RDM services in the field of plasma medicine
- Discussion about previous work and intended next steps in the QPTDat project to ensure the highest possible benefit for the community

Agenda

Day 1

- 13:15 [Welcome and introduction](#)
- 13:30 [Ontologies and Knowledge Graphs for FAIR Research Data Management](#)
- 14:05 [Methods and data in plasma/bio-medicine](#)
- 14:40 [Tool pipelining and data analysis with KNIME](#)

- 15:30 [FAIR data management in biomedicine](#)
- 16:05 [Contributed use cases and discussion](#)
 - 16:10 [Plasma effects on endoglucanase enzyme](#)
 - 16:15 [Plasma and nanotechnology for oncotherapy](#)
 - 16:20 [Modelling and modulating the concentration of reactive oxygen and nitrogen species \(RONS\) in different mixture at atmospheric pressure non thermal plasma KINPen/Jets](#)
 - 16:25 [Plasma agriculture](#)
 - 16:30 [FRBRization of knowledge graphs about research data](#)
 - 16:35 [Summary](#)

Day 2

- 13:15 [Innovation-Platform MaterialDigital](#)
- 13:50 [Research data management within the research department plasmas with complex interactions, especially plasma medicine research groups at the Ruhr-University Bochum](#)
- 14:25 [Community based FAIR Research Data for Plasma Medicine](#)
- 14:40 [Use cases to integrate sharing of plasma technology research data with blockchain technology](#)
- 14:55 [Application of ontology and blockchain for FAIR research data in plasma medicine](#)

- 15:30 [Discussion groups](#)
- 16:30 [Summary](#)

Day 1

Welcome and introduction

Markus Becker, Leibniz Institute for Plasma Science and Technology, Greifswald

[\[Slides\]](#)

Ontologies and Knowledge Graphs for FAIR Research Data Management

Harald Sack, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe

Summary

Harald Sack gave a first overview over ontologies and knowledge graphs by showing examples and different definitions of the terms. He presented the benefits of ontologies as a common language of knowledge: How they help with searching and categorizing data as well as applying the FAIR data principles. He also explained how an ontology could be implemented on the level of applications with knowledge graphs and connected to already existing ones to build an overall basis of knowledge in the long term perspective.

[\[Slides\]](#)

Methods and data in plasma/bio-medicine

Sander Bekeschus, Leibniz Institute for Plasma Science and Technology, Greifswald

Summary

In his talk, Sander Bekeschus reviewed chances and difficulties of data management in the field of plasma medicine. He showed how the complexity of biomedicine increases the closer one gets to real patients and how the interaction of plasma and biomedicine leads to a multiplied complexity. Based on different examples, he showed where metadata could lead to high reproducibility, but also where limits could be and how essential it is to track the whole process from raw to analyzed data. He especially hinted that even single products of the setup could have a major impact on results and need to be tracked. He went on presenting methods that already contain metadata files or where databases already exist.

[\[Contact\]](#)

Tool pipelining and data analysis with KNIME

Nick Plathe, Leibniz Institute for Plasma Science and Technology, Greifswald

Summary

Nick Plathe introduced the tool KNIME for data analysis and gave details of an implemented data workflow. KNIME is a free software (client version) with graphical programming. He showed that this tool is suitable to give access to data driven research and how it can be used to extract and combine information from external binaries and databases. He also presented how KNIME could be used to display this information and semi automate workflows.

[\[Slides\]](#)

FAIR data management in biomedicine

Dagmar Waltemath, Medical Informatics, University Medicine Greifswald

Summary

Dagmar Waltemath showed for the related topic of biomedicine how data management can be implemented. She emphasized the importance of standardizing visualization, model encoding and simulation encoding and linking all of them. She reported that it took around 20 years to get from mere paper publishing to reproducible and standardized experiments. To achieve this, an archive bundling all data needed to reproduce experiments was developed. She described this process as community driven, professionalized over time (umbrella organization) and by being in contact with consortia, funders and journals. She underlined that data management needs to be advertised in and developed with the community. She finished by showing how useful this knowledge is in the current topic of Covid-19.

[\[Slides\]](#)

Contributed use cases and discussion

Plasma effects on endoglucanase enzyme

Zeinab Kabarkouhi, Shahid Beheshti University, Tehran

[\[Slides\]](#)

Plasma and nanotechnology for oncotherapy

Milad Rasouli, Kharazmi University, Institute for plasma research, Tehran

[\[Slides\]](#)

Modelling and modulating the concentration of reactive oxygen and nitrogen species (RONS) in different mixture at atmospheric pressure non thermal plasma KINPen/Jets

Junaid Khan, Comsats University Islamabad

Plasma agriculture

Robert Wagner, Leibniz Institute for Plasma Science and Technology, Greifswald

[\[Slides\]](#)

FRBRization of knowledge graphs about research data

Ingo Frank, Leibniz Institute for East and Southeast European Studies, Regensburg

[\[Slides\]](#)

Summary

In the first four lightning talks different use cases from plasma science were introduced and an overview over the kind of data that results were given. The last lightning talk given by Ingo Frank introduced an approach to FRBRize knowledge graphs about research data for versioning of datasets, extended provenance information, and improved data citation.

Day 2

Innovation-Platform MaterialDigital: Towards a Deep-rooted Interaction of Ontology and Workflow Tools in Materials Science and Engineering

Markus Niebel, Fraunhofer Institute for Mechanics of Materials IWM, Freiburg

Summary

In his talk, Markus Niebel introduced the diverse field of materials science and engineering and approaches and aims in data management. The goal is to share knowledge between different fields and companies with the same research topics and at different points of manufacturing chains by means of the Innovation-Platform MaterialDigital. He pointed out how an optimization of material usage brings benefits concerning sustainability, financial profit as well as resource usage. He further showed that the description of materials needs data, metadata and interpretation tools. He reported on efforts on balancing this with a broad approach in the core ontology and the use of blobs to simplify the description of very complex structures. On the technical side, he showed tools to communicate between experts

of the field and data managers as well as a central database for linking of decentralised infrastructures.

[\[Slides\]](#)

Research data management within the research department plasmas with complex interactions, especially plasma medicine research groups at the Ruhr-University Bochum

Marina Prenzel, Ruhr-Universität Bochum

Summary

This talk by Marina Prenzel gave an overview about research data and data management in plasma medicine at Ruhr-University Bochum. Marina Prenzel showed internal activities as well as collaborative efforts together with INP Greifswald and within the framework of the NFDI4Phys consortium. She gave examples of different research groups and their data both in simulation and laboratory experiments. She also explained where metadata already exists and which challenges and requirements are seen in those groups. All groups communicate the need for IT-support and see linkage between different data, data literacy and time management as common challenges.

[\[Slides\]](#)

Preview of an ontology and blockchain use cases for research data management in plasma medicine and plasma technology

Community based FAIR Research Data for Plasma Medicine

Tabea Tietz, Fabian Hoppe, Harald Sack, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe

Summary

Tabea Tietz and Fabian Hoppe held the first of three talks about the project QPTDat and the work done in this project concerning data management in plasma medicine. They focused on the aspects FAIR data and community based data management. They also described the process of developing an ontology in collaboration between data management experts and experts in the scientific field. Furthermore, they showed how the starting point of the ontology looks like and how it can be used and extended by the community via SPARQL and Wikibase and what specifics an ontology needs to have in this setting.

[\[Slides\]](#)

Use cases to integrate sharing of plasma technology research data with blockchain technology

Simon Tschirner, Hochschule für Angewandte Wissenschaften Hamburg

Summary

In the second part of the introduction of the QPTDat project, Simon Tschirner talked about the benefits of blockchain in data management. He first explained what a blockchain is and how it will be used within this project to check the integrity of data. He explained how data ownership can be secured and validated and how this can help in a reputation management for published data. He also introduced Bloxberg and the benefits of this specific Blockchain network for scientific data as an authority based and decentralized system. He finished by showing use cases for the blockchain within different stages of research data management.

[\[Slides\]](#)

Application of ontology and blockchain for FAIR research data in plasma medicine

Laura Vilardell Scholten, Markus Becker, Leibniz Institute for Plasma Science and Technology, Greifswald

Summary

This talk was held by Markus Becker and dealt with the points of contact that researchers with ontologies, knowledge graphs and the blockchain have or might have in the future. He first showed the complexity of the field of science and how an ontology can help to tackle this complexity with respect to research data management. He went on showing examples on how this could look starting in the lab using an electronic laboratory notebook, saving data and metadata alongside and going to data publishing, searching for data by means of metadata and reusing datasets. He also emphasized how the blockchain helps with improving trust in the data and reputation monitoring.

[\[Slides\]](#)

Discussion groups

G1: Metadata and quality criteria

Markus Becker

Introduction

Discussion about the modular and process-oriented approach (description of methods, devices, objects, ...) pursued in QPTDat for data documentation and formulation of quality criteria. Does this fit with your requirements? Which metadata is considered to be important? What is already recorded at present?

[\[Slides\]](#)

Conclusion

- General metadata structure and ontology based approach seems to be practical.
- Further application within repositories and usage for documentation of datasets also from simulation studies would be the next step (“stress test”).
- Specific requirements and necessary metadata fields will evolve from application of the first versions of the metadata modules, e.g. addition of more parameters according to biology etc.
- Interdisciplinary research poses additional challenges → how to connect the plasma metadata with metadata from other fields (biology, chemistry etc.)? E.g. protocols are important in biology.
- A further problem could be that often the results (data) depend very sensitively on certain parameters, e.g. the geometry of the plasma source. Achievement of reproducibility will be very challenging in this case even if the geometry is described by metadata.
- Community wide quality criteria are a must and the possibility of centralized quality checks according to the ontology seems to be effective. But the fact that some requirements are very specific to local setups presents a major challenge.

G2: Collection and publication of research (meta)data

Fabian Hoppe

Introduction

Discussion about the publication of data and metadata in repository and/or knowledge graph. Wikibase, which is the basis of Wikidata (<https://www.wikidata.org>), is considered as a community tool to establish and maintain knowledge graphs.

[\[Slides\]](#)

Conclusion

- Often the publication of data and metadata is seen as a desirable/optional step in addition to a paper. Therefore, the publication of data has to be easy to understand and fast, because scientists cannot invest significant time due to the pressure to publish papers instead.
- A wiki based approach as a collaboration tool is practical, because many research groups already work frequently with Wikis, which reduces the “barriers to entry”.

- The basic process of publishing data on the Wikibase platform seems to be easy to understand. Nevertheless, many scientists are interested in bulk-upload options to make the publishing process more efficient.

G3: What makes research data FAIR?

Tabea Tietz

Introduction

Discussion about workflows of generating research (meta)data in the community. At what point in the research process does research (meta)data have to be collected? What does it take for research (meta)data to be findable, reproducible and reusable? What are the requirements?

[\[Slides\]](#)

G4: Data certification and reputation monitoring

Simon Tschirner

Introduction

Discussion about the relevance of data certification, validation of data provenance, and reputation monitoring. At what point should data be certified? Whom would you trust for certification? What happens if data cannot be validated? Which factors should influence reputation? How important is citation and reputation related to research data?

Conclusion

The discussion has shown that there is certainly an interest in having the possibility to prove authorship of research data. However, it also turned out that there is still a high demand for further explanation of the used Blockchain technology. This leads to the conclusion that an important aspect to explore would be: “How to communicate appropriate details about the backend technology to the users via the user interface?”.

Summary

In summary, research data management experts provided an overview over existing procedures and ideas for the provision and usage of quality assured research data, e.g. as a basis for data-driven science. On the other side, researchers in the field of plasma medicine and plasma technology introduced different use cases and pointed out specific demands and possible challenges. On this basis, approaches for the further development of an ontology, metadata standards, and quality criteria supporting a simplified publication and re-use of data in the field of plasma medicine were worked out during the group discussions. The collected ideas will be incorporated into the services to be developed in the framework of QPTDat and thus be made available to the community.

By means of short surveys the participants reported that the workshop was according to their expectations and helped to understand the research data management concepts and the meaning of the FAIR data principles. Most participants of the survey already see a benefit in making their data FAIR. Future workshops should give more time to the discussions, the mix of topics from research data management and plasma science was appreciated.